Frank Chlebana
Mar 26 2002

DAQ

The CDF Data Acquisition consists of the infrastructure, electronics and software used to collect data, calibrate detectors, monitor and configure the electronics.

The Front End (FE) electronics consists of a combination of custom built modules, designed and built by many different institutes and universities from across the world, and commercial off the shelf hardware.

Security Issues

As part of your job you are required to use computers that are designated as part of the CDF Critical System.

It is your responsibility to help minimize the possibility of loading software that has an adverse effect of the performance of these systems (e.g. software viruses, hacker kits). The most frequent incident we've had in the critical system involves PC viruses loaded on windows based PCs.

A common way that a virus is loaded on windows based PCs is through email attachments. From a critical system computer you should not open email that is not needed for the critical system operations or is unexpected. Be aware that many new viruses can spread via open Windows file shares or downloads from websites (so beware mail saying "click here").

The windows based PCs are to be used for Slow controls functions only. Email, web, and any other general purpose access should be done on some other system. There are linux based PCs that can be used for this purpose.

Do not install any software on the CDF Critical System PCs. If you feel that there is a software package which would be useful please contact one of the system administrators to arrange for the package to be installed.

Your other major responsibility is to avoid letting non-authorized persons gain access to the critical systems through your means. Keep your password secure, difficult to guess and secret. Have your account disabled if you won't need it for months. Watch for indications that your account has been used by someone else. Do not leave your session unattended for extended periods of time.

In the past, systems on site have been compromised and these compromised systems have initiated attacks on offsite networks...

Fermilab is one of the more open labs. If it is perceived, by the people who fund us, that we are unable to manage our computers in such an open fashion we will be pressured to tighten security up...

The online network is split into a "lower half" (trigger room) and "upper half" (control room).

The lower half is used for development and monitoring tasks that are not critical for taking data while the upper half is used to support data taking.

General use PCs.

b0dap06.fnal.gov	b0dap01.fnal.gov
b0dap12.fnal.gov	b0dap16.fnal.gov
b0dap18.fnal.gov	b0dap21.fnal.gov
b0dap19.fnal.gov	b0dap26.fnal.gov
b0dap20.fnal.gov	

From the CDF offline network (trailers) onsite you can log in directly to the trigger room PCs.

From offsite you have to go through one of the designated gateway node, b0dap30 or b0dau30 or log into a trailer PC then into the lower half of the online network.

Kerberos and You or... How to Get Around

In general one should NOT type in a password over the network. Once logged into a local machine you can forward your credentials to the next machine using *ssh* or *rlogin -F*

```
> ssh hostname
> rlogin -F hostname
```

ssh configures your x window environment so you do not have to set the display variable and this is the simplest way of logging into another node.

⇒ User is on a group account cdfdaq and wants to go to another PC with the same group account.

```
b0dap50_cdfdaq> kticket
b0dap50_cdfdaq> ssh nodename
```

⇒ User is logged onto a group account and wants to get a ticket for their own account.

```
b0dap50_cdfdaq> kinit username
b0dap50_cdfdaq> ssh nodename
```

This ticket is valid only from the window you issued the command. You will be logged into the remote node as the user you specified in the kinit command.

⇒ Can display a X window originating from the remote machine.

On the remote PC (fcdfsgi2)

```
fcdfsgi2_username> xterm  (to check if X works)
```

When work is done.

```
fcdfsgi2_username> exit
b0dap50_username> kdestroy
```

One should enter "kdestroy" in the same window where you issued done "kinit", to destroy the ticket.

⇒ What controls access for a user to an account?

Users can log into an account provided that they are in the .k5login file. This file is owned by root.

→ The first three people in the .k5login file are responsible for the group account.

To get around, use kerberos rlogin or kerberos telnet or kerberos ssh with kerberos ticket.

To check current ticket

```
b0dap50_cdfdaq> klist
```

To check a default rlogin (telnet, ssh)

```
b0dap50_cdfdaq> which rlogin (telent, ssh)
/usr/krb5/bin/telnet
```

To list your current tickets

```
> klist
```

If you have problems (in getting from one machine to another)... let us know.

Overview of the DAQ Components

Trigger Supervisor and Crosspoints

Receives trigger decisions from the L1 and L2 global trigger systems and distributes trigger signals though the trigger cross points to the DAQ components.

Front End and Trigger VME Crates

Reads out, formats and transfers the data.

Event Builder

Assembles data fragments from the many FE crates into one block.

Level 3 Trigger

Formats the data into the final data format (root). Have access to the entire event record and can run offline code to make trigger decisions.

Consumer Server/Logger

Receives data from L3 and writes data to disk in several different data streams based on L3 trigger results. Also distributes events to the consumers.

Run Control

Coordinates data taking and detector calibrations.

DAQ Monitoring

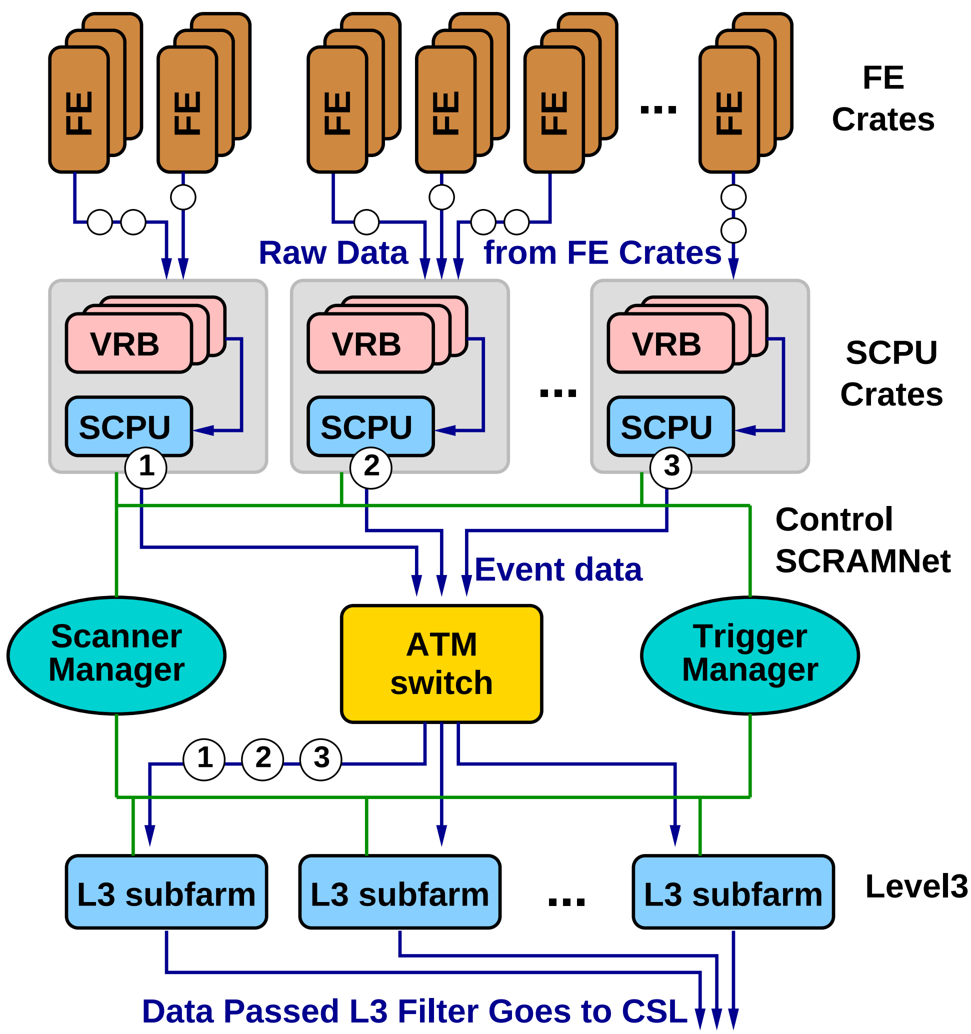
Monitors DAQ performance.

Data Monitoring (Consumers)

Monitors the data quality.

The data is self describing. The VME Readout Controller (VRC) forms a mini-bank by attaching a header, which identifies the data type and the block number, to the data block.

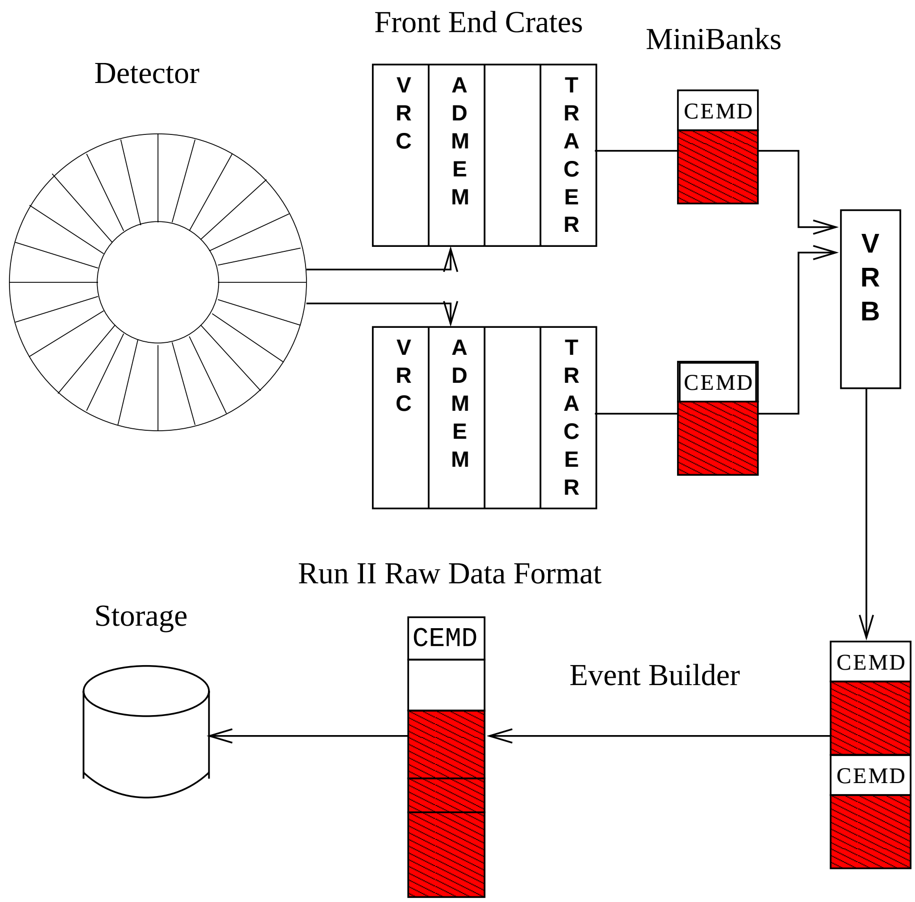
This mini-bank is then transferred to one of 15 VME Readout Buffer (VRB) crates. Several VRBs in a crate and each VRB can receive data from up to ten FE crates.



The event fragments are then sent through the ATM switch to a converter node which distributes the data to the processor nodes of L3.

The reformatter code assembles the minibanks into the final event format making it available to the L3 analysis code.

→ Events having corrupted fragments are rejected by the reformatter.



At L3 the reconstruction can add reconstructed objects to the event record.

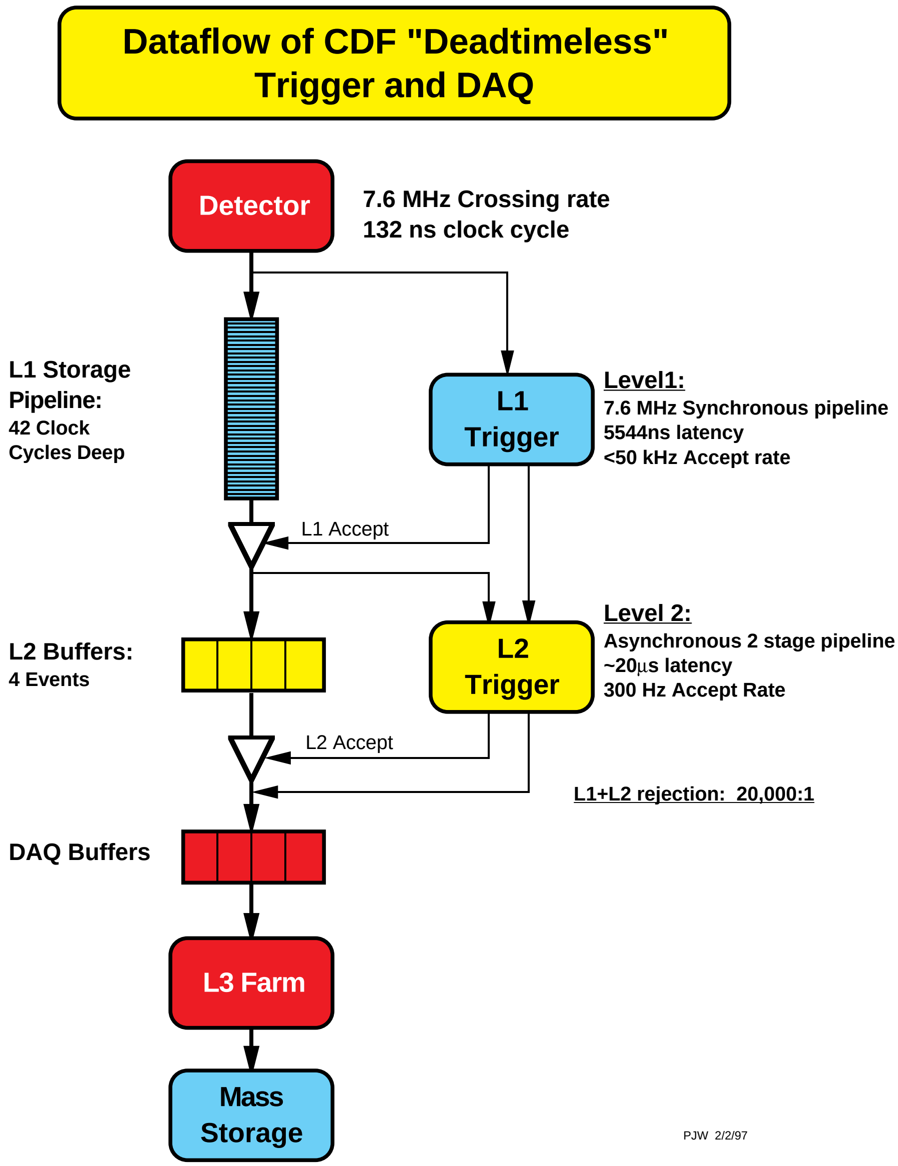
Events passing the L3 trigger are sent to the Consumer Server Logger and a fraction of the events are distributed to monitoring consumers.

Events are transferred to Feynmann Computing Center (FCC) for storage on tape.

Users access the data through the Data Handling System.

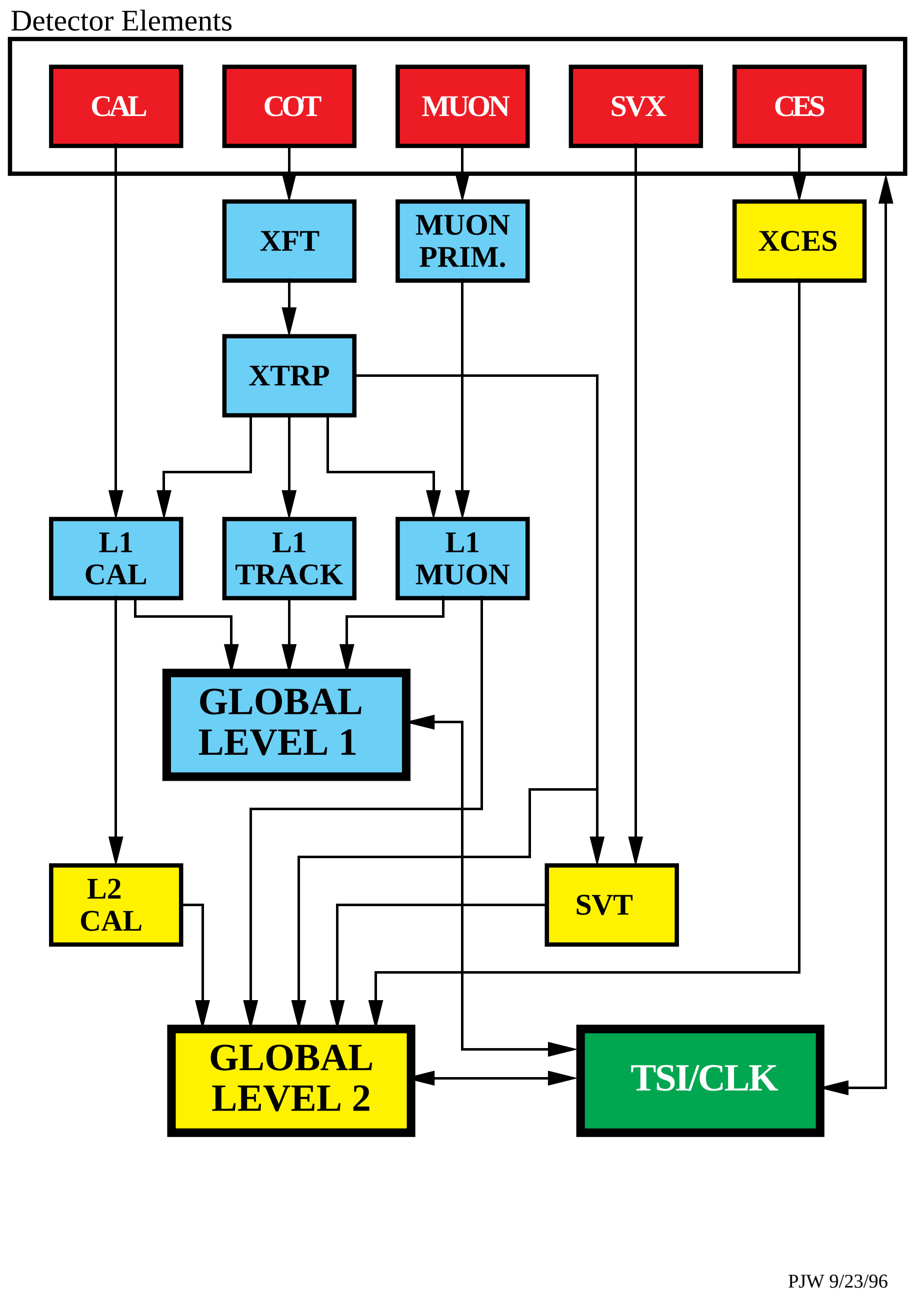
Trigger System

The trigger system is used to select an event rate of 75 Hz from the 7.6 MHz (132 ns crossing) beam crossing rate.



Done in three stages: Beam crossing rate of 7.6 MHz (for 132 ns crossing) is reduced to $\leq$ 50 KHz, by the L1 trigger, reduced to 300 Hz by L2 and finally to about 75 Hz by L3.

RUN II TRIGGER SYSTEM

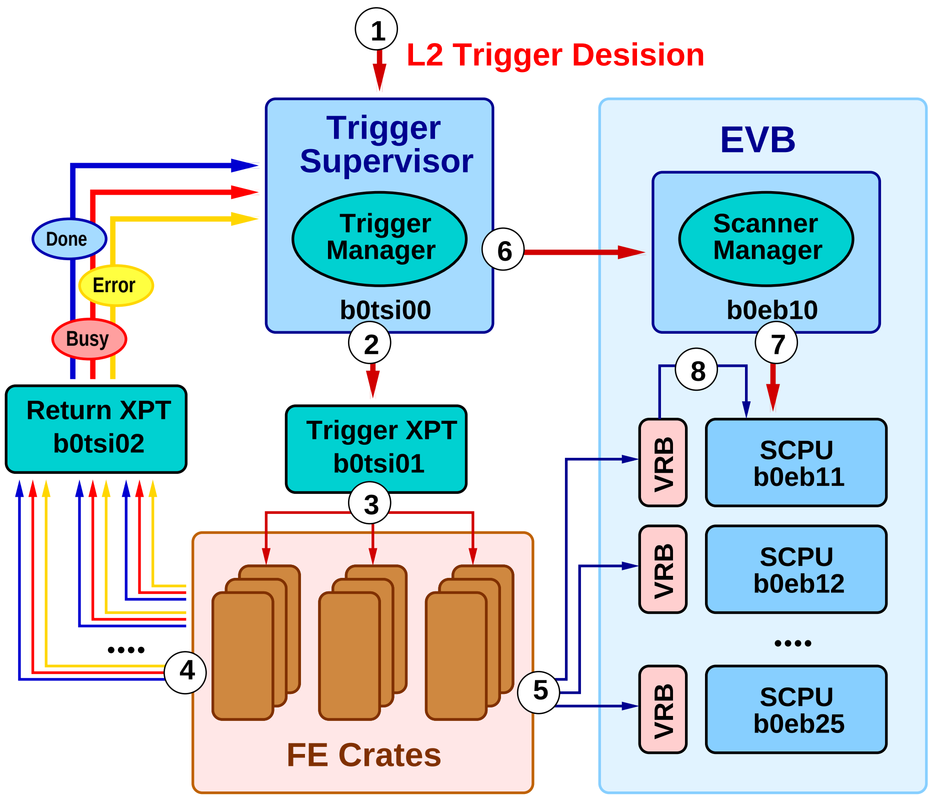


Each stage has access to more complete data and has more time to process the event. At L3 you have the final complete event and can run sophisticated offline algorithms.

So far we have been running with a L1 trigger at rates up to 400 Hz, L2 has been passing all events and we are rejecting jet events at L3 below threshold.

The L2 Trigger sends the trigger decision to the Trigger Supervisor (TS).

The L2 trigger decision is sent to the FE crates through the trigger cross point.



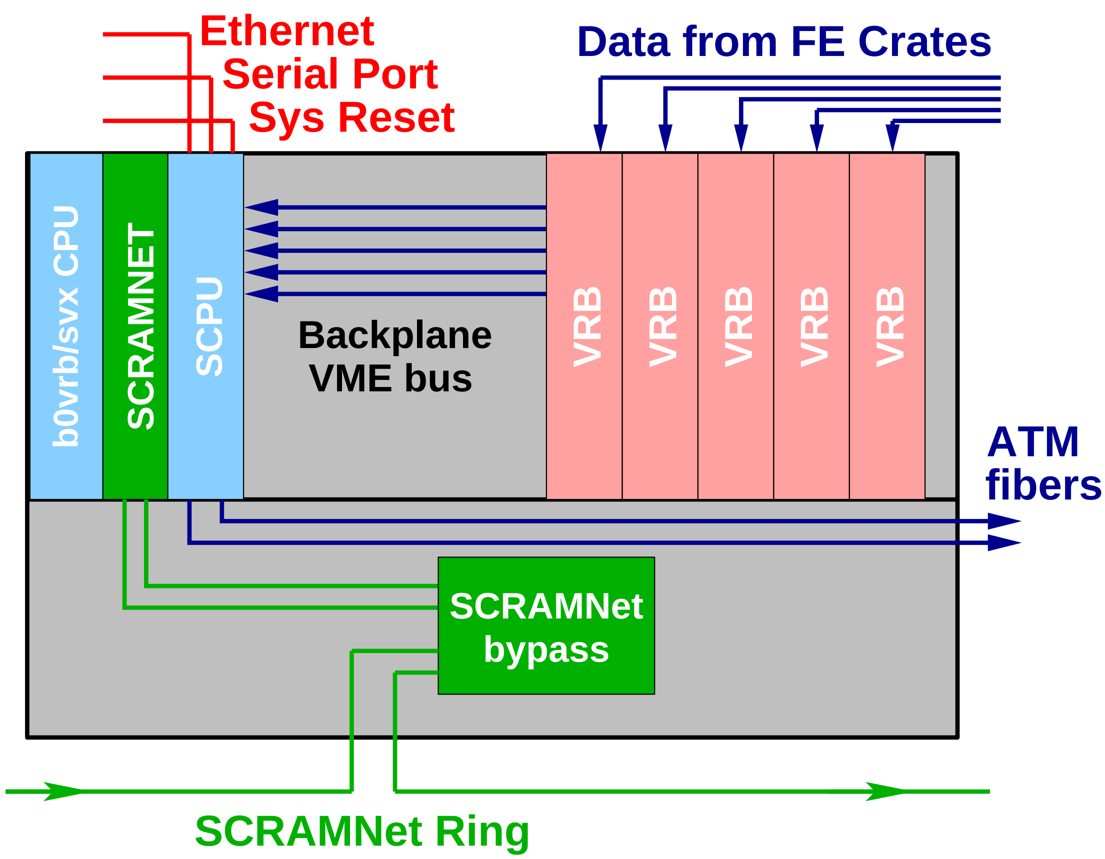
When the trigger decision is received in the FE crate, the VME Readout Controller (VRC) sends back a DONE signal to the TS via the return cross point indicating that it is ready to receive the next trigger decision.

Data is read out from the FE cards (TDCs, ADMEMs...) formatted and sent via the TRACER to the VME Readout Buffers VRBs.

If there is not enough space to write out the event to the VRB a BUSY signal is sent back to the TS so that the TS does not issue another trigger which leads to busy deadtime.

→ If the busy is not deasserted in time we can get a Busy Timeout causing the run to halt.

Each VRB can receive data from up to 10 different front end crates.



Data is readout of the VRBs by the Scanner CPU (SCPU) and sent through the ATM switch to the L3 farm.

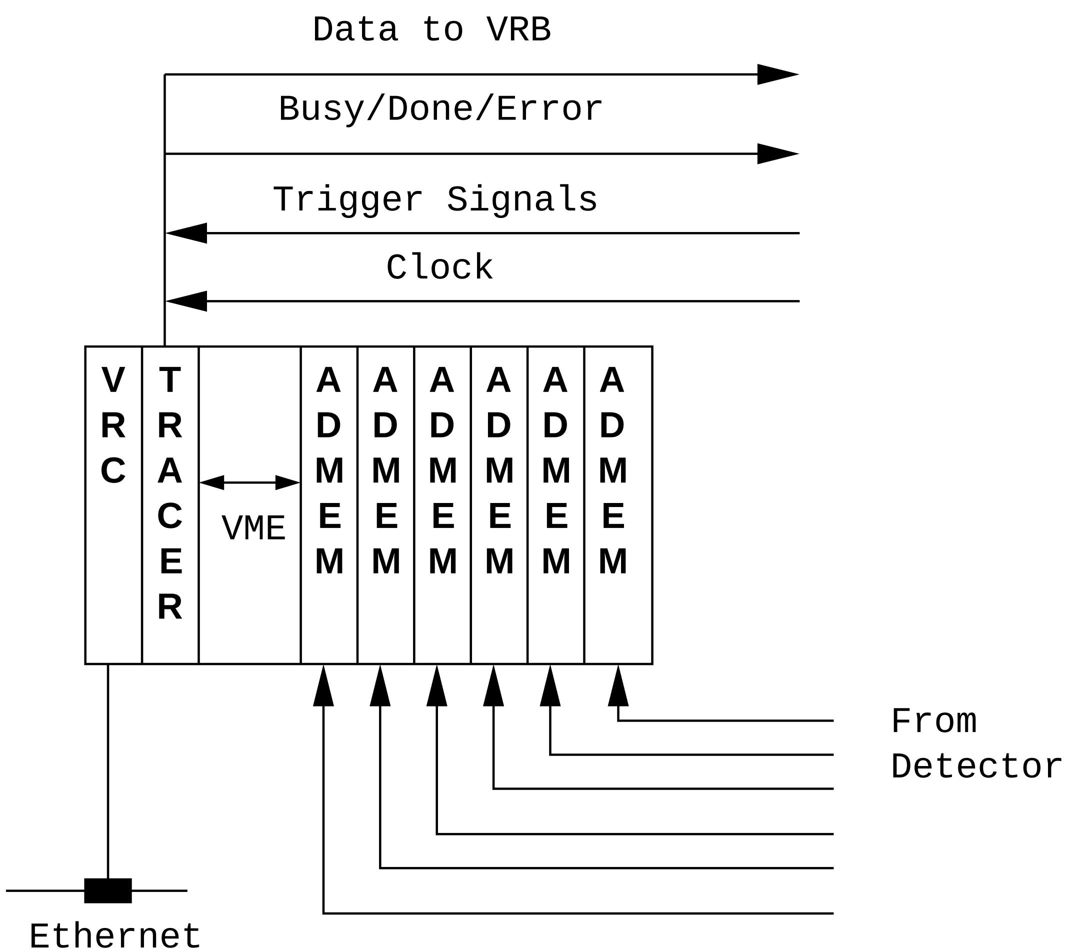
The SCPU communicates with the TS via SCRAMNET.

Typical Front End/Trigger Crate

Front End (FE) and Trigger electronics are housed in VME crates, an industry standard backplane (aka a crate) into which compliant cards can be plugged into.

`http://www-esd.fnal.gov/esd/catalog/vmedir.htm`

A typical crate will have a VME Readout Controller (VRC). Usually a Motorola MVME 2301 with a Power PC 603 CPU running VxWorks, a TRACER used to fan out trigger and clock signals to the VME backplane and to transport data out of the crate, and the FE electronics.



VxWorks is a real-time operating system having a fast interrupt response and network connection.

The front end crates have node names such as b0pcal00 etc. You can log in to the crate to check the status but this is not usually necessary during normal running. It is useful for tracking down problems.

There are about 120 FE/Trigger crates, about half of which are mounted on the detector and are not accessible during collisions.

Crates will also have a TRACER. This card receives the clock and trigger signals and distributes them on the VME back plane for the other modules to pick up.

Returns control signals back to the TS and provides a data path to the VME Readout Buffers.

If the network path (ethernet) to the VRC is not available you can try accessing it via minicom (serial line) from b0dap10.

You can also reset the crate using the system reset lines in order to force it to reboot. Normally this should be tried only as a last resort.

Silicon System

The silicon system has a different architecture that the rest of the DAQ.

Trigger information from the TS is sent to the Silicon Readout Controller at which point it is distributed to the rest of the silicon system.

The processor in the Silicon crates is used for configuration and monitoring - not used for readout.

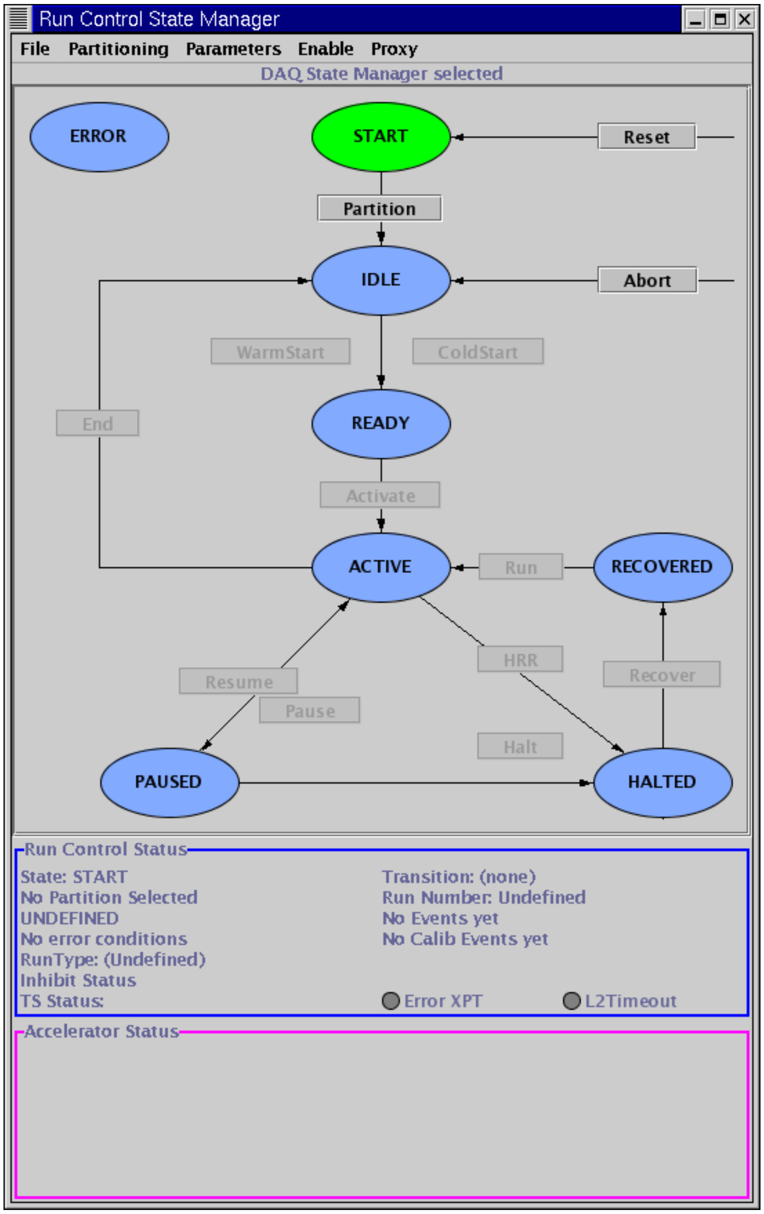
Data is transferred from the front end crates through the FIB crates to the VRB crates.

More details in a separate talk...

Run Control

Coordinates the configuration, starting and stopping of runs.

Written in Java, uses a Graphical User Interface showing a state transition diagram to control the many distributed clients.



Can group together clients into a *partition*. Can run with up to eight hardware partitions simultaneous.

Allows inclusion or exclusion of individual cards or crates, masking of bad channels...

Uses the commercial message passing software package smartsockets.

Uses DaqMsg (layered on top of smartsockets) to provide automatic code generation to conveniently pack and unpack data structure (messages...).

Machine independent communication... clients written in Java running under Linux communicating with clients written in C running under VxWorks.

Clients subscribe to a subject and *configuration and control* messages are broadcast to all clients who are subscribed to a particular message.

Subjects have the syntax

/partition-0/frontEnd/ccal/00

Can use wildcarding to broadcast message to all clients of a certain type. Used for "run sequencing", to bring a group of clients through a transition before a second group of clients.

All communication goes through the *rtserver*.

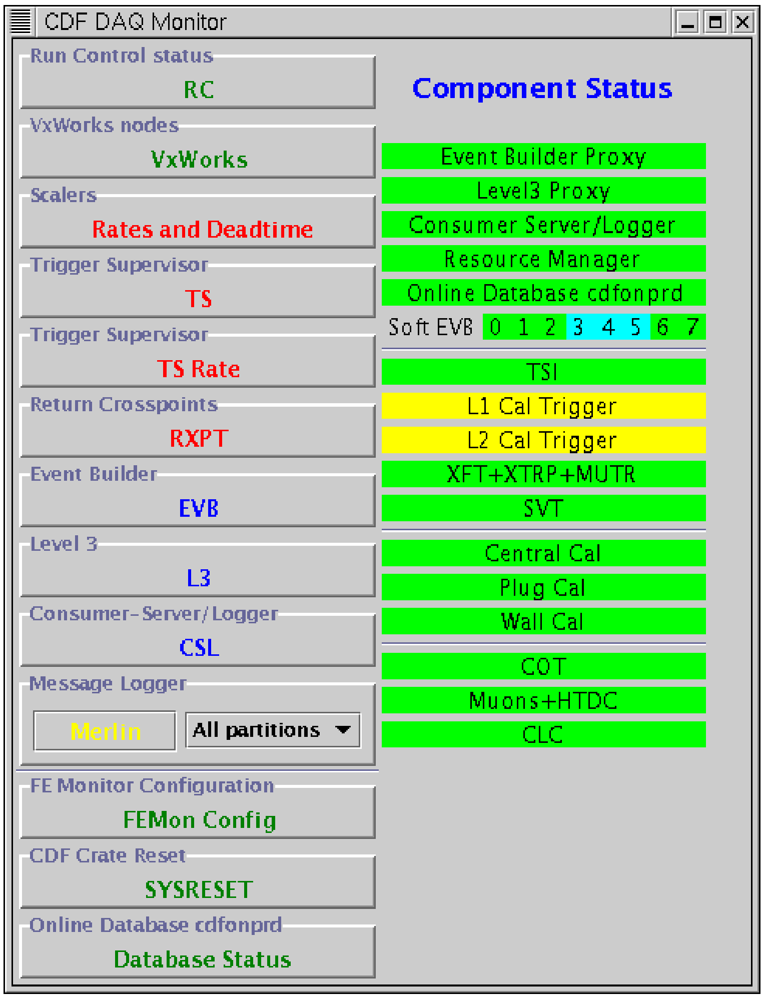
If the RC GUI crashes the run could still be going... and you can try reattaching a new RC to the current run.

More details in a separate talk...

DAQ Monitoring

There are a number of programs that can be used to monitor the performance of the DAQ. Can be launched from a main control panel...

```
> setup fer
> daqmon
```



These monitors are mostly used to check that data is flowing through the system.

The quality of the data is checked by the consumers.

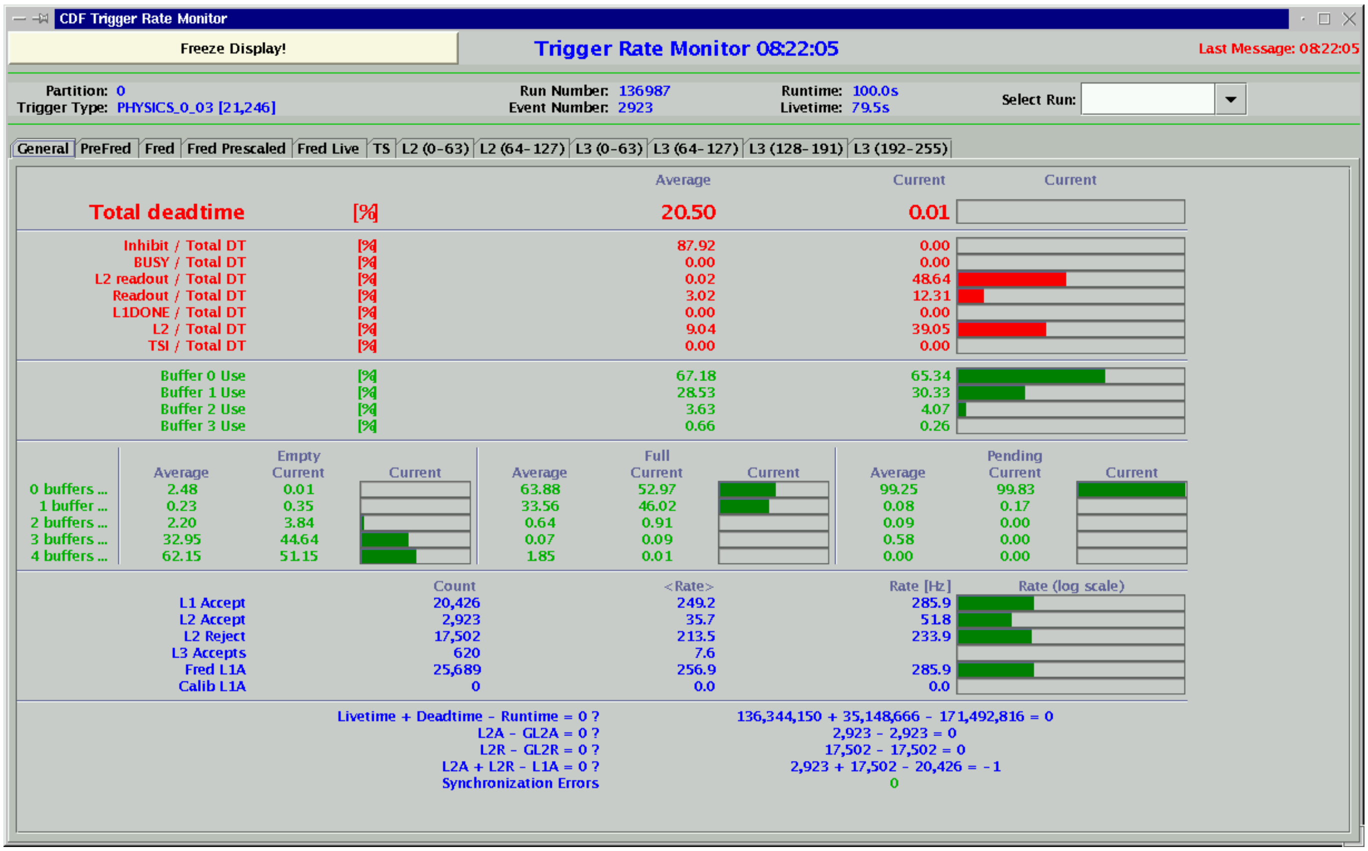
The more popular DAQ monitors are:

L3

Gives an overview of how L3 is working

Rates and Deadtime

General display...

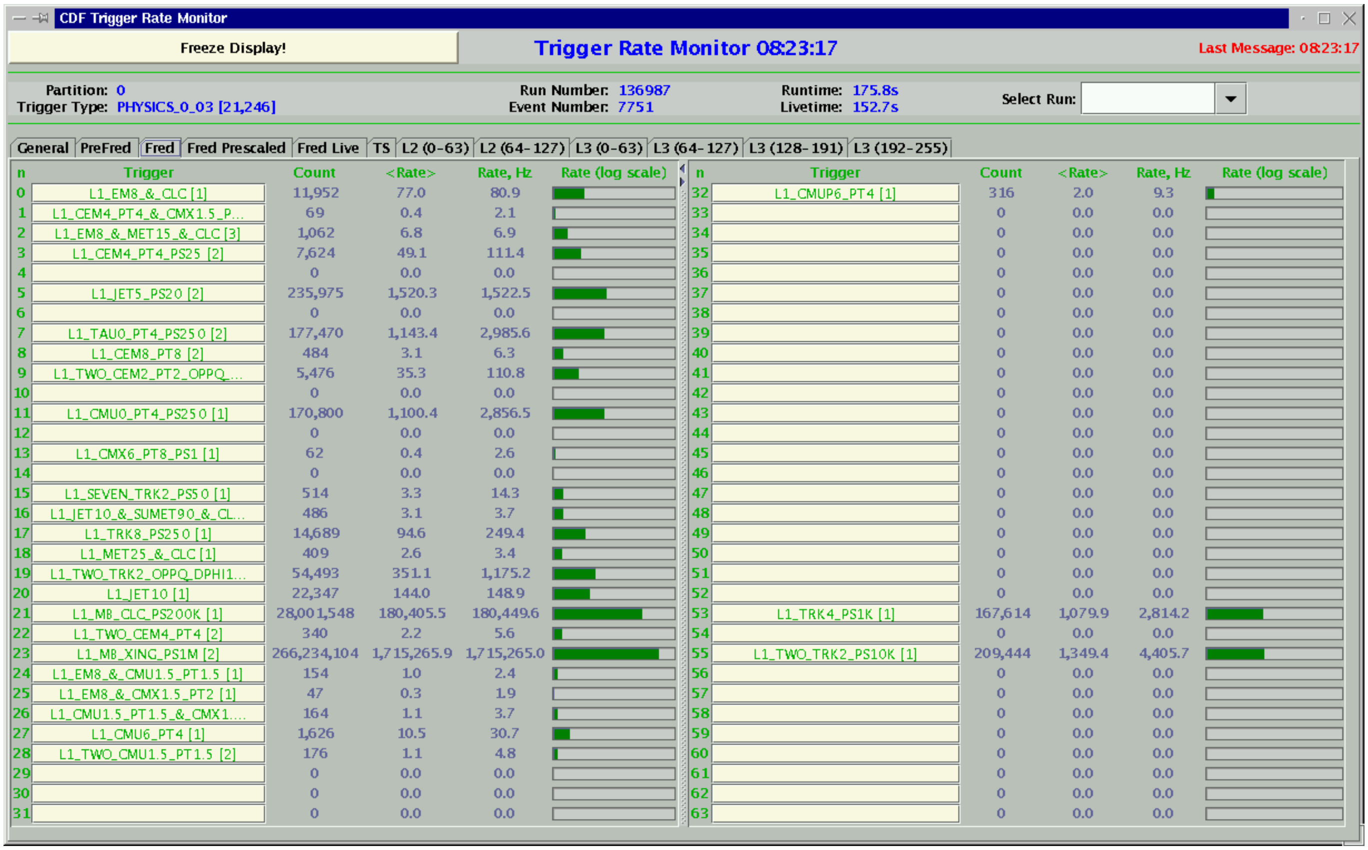


Useful to check that the trigger is properly functioning.

Shows the total system deadtime and how it is allocated.

Several tabed panels are available to give you a detailed look at the rates for each trigger at L1, L2 and L3.

Use one of the tabs to look at specific trigger rates in more detail...



n	Trigger	Count	<Rate>	Rate, Hz	Rate (log scale)
0	L1_EM8_&_CLC [1]	11,952	77.0	80.9	
1	L1_CEM4_PT4_&_CMX1.5_P...	69	0.4	2.1	
2	L1_EM8_&_MET15_&_CLC [3]	1,062	6.8	6.9	
3	L1_CEM4_PT4_PS25 [2]	7,624	49.1	111.4	
4		0	0.0	0.0	
5	L1_JET5_PS20 [2]	235,975	1,520.3	1,522.5	
6		0	0.0	0.0	
7	L1_TAU0_PT4_PS250 [2]	177,470	1,143.4	2,985.6	
8	L1_CEM8_PT8 [2]	484	3.1	6.3	
9	L1_TWO_CEM2_PT2_OPPQ_...	5,476	35.3	110.8	
10		0	0.0	0.0	
11	L1_CMU0_PT4_PS250 [1]	170,800	1,100.4	2,856.5	
12		0	0.0	0.0	
13	L1_CMX6_PT8_PS1 [1]	62	0.4	2.6	
14		0	0.0	0.0	
15	L1_SEVEN_TRK2_PS50 [1]	514	3.3	14.3	
16	L1_JET10_&_SUMET90_&_CL...	486	3.1	3.7	
17	L1_TRK8_PS250 [1]	14,689	94.6	249.4	
18	L1_MET25_&_CLC [1]	409	2.6	3.4	
19	L1_TWO_TRK2_OPPQ_DPHI1...	54,493	351.1	1,175.2	
20	L1_JET10 [1]	22,347	144.0	148.9	
21	L1_MB_CLC_PS200K [1]	28,001,548	180,405.5	180,449.6	
22	L1_TWO_CEM4_PT4 [2]	340	2.2	5.6	
23	L1_MB_XING_PS1M [2]	266,234,104	1,715,265.9	1,715,265.0	
24	L1_EM8_&_CMU1.5_PT1.5 [1]	154	1.0	2.4	
25	L1_EM8_&_CMX1.5_PT2 [1]	47	0.3	1.9	
26	L1_CMU1.5_PT1.5_&_CMX1....	164	1.1	3.7	
27	L1_CMU6_PT4 [1]	1,626	10.5	30.7	
28	L1_TWO_CMU1.5_PT1.5 [2]	176	1.1	4.8	
29		0	0.0	0.0	
30		0	0.0	0.0	
31		0	0.0	0.0	

n	Trigger	Count	<Rate>	Rate, Hz	Rate (log scale)
32	L1_CMUP6_PT4 [1]	316	2.0	9.3	
33		0	0.0	0.0	
34		0	0.0	0.0	
35		0	0.0	0.0	
36		0	0.0	0.0	
37		0	0.0	0.0	
38		0	0.0	0.0	
39		0	0.0	0.0	
40		0	0.0	0.0	
41		0	0.0	0.0	
42		0	0.0	0.0	
43		0	0.0	0.0	
44		0	0.0	0.0	
45		0	0.0	0.0	
46		0	0.0	0.0	
47		0	0.0	0.0	
48		0	0.0	0.0	
49		0	0.0	0.0	
50		0	0.0	0.0	
51		0	0.0	0.0	
52		0	0.0	0.0	
53	L1_TRK4_PS1K [1]	167,614	1,079.9	2,814.2	
54		0	0.0	0.0	
55	L1_TWO_TRK2_PS10K [1]	209,444	1,349.4	4,405.7	
56		0	0.0	0.0	
57		0	0.0	0.0	
58		0	0.0	0.0	
59		0	0.0	0.0	
60		0	0.0	0.0	
61		0	0.0	0.0	
62		0	0.0	0.0	
63		0	0.0	0.0	

→ Normally the physics triggers are defined so that the dead time is less that about 5%.

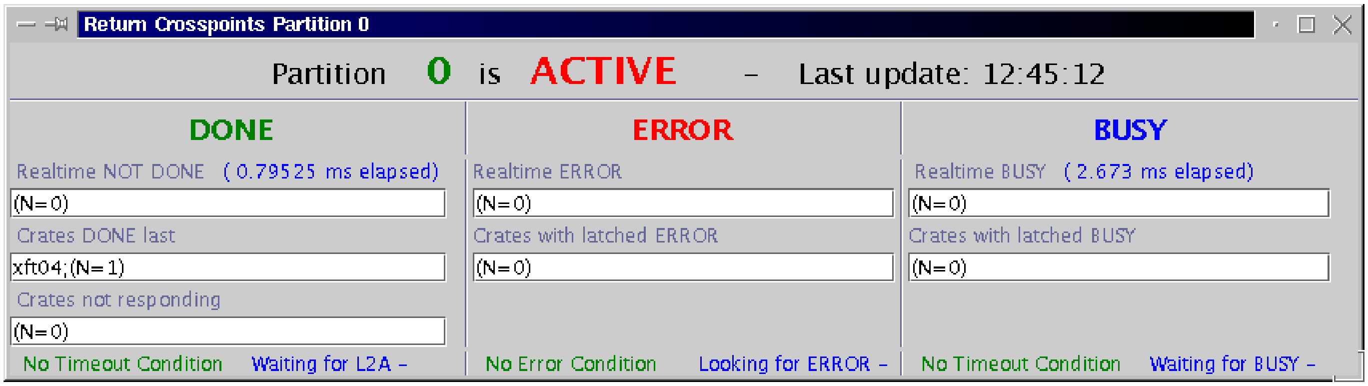
If you have a run going with high dead time this indicates a problem which needs to be fixed.

You can sometimes see which trigger is firing at a high rate from the “Rates and Deadtime” GUI.

RXTP

Shows which client was the last to return DONE, BUSY and ERROR and shows the time it took.

Useful to identify which crate is contributing to the deadtime.



In this example the XFT04 crate was the last to return a Done, and it was set 0.79 ms after the L2 decision was received.

→ Taking a long time to set the Done can result in "Readout Deadtime".

Typically a crate should set the done within 1 ms, but there are a few crates which can take longer.

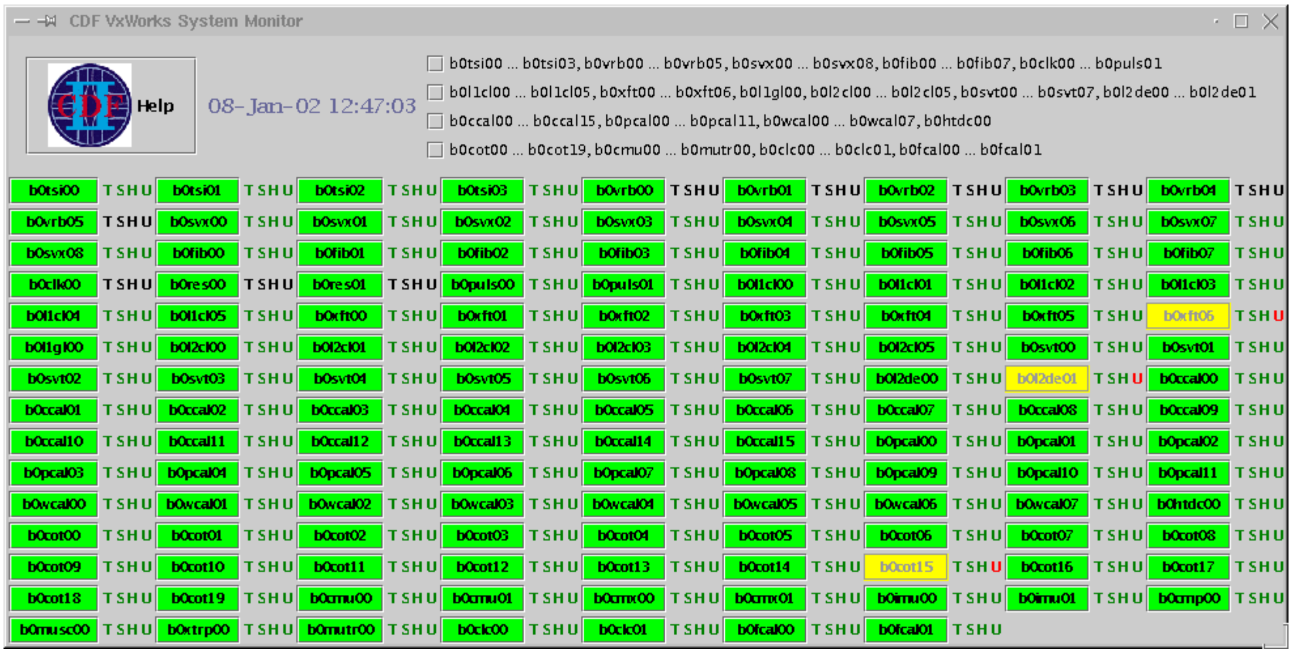
There is also a consumer (daqmon) that histograms the readout times and event sizes, more later...

CSL

Shows the status of the CSL, logging rates, partitions etc...

VxWorks Monitor

The VxWorks monitor gives an overview of the status of the Front End crates in the system.



Each button corresponds to one of the front end crates.

Green indicates that the process is OK

Yellow indicates that the crate is not updating information

Red indicates that problem with the crate

Useful to see if a front end crate has crashed...

Error Handler

Error messages from the different clients are sent to the Error Handler, which displays the message on the screen and also logs the error messages.

After setting up the fer package, (setup fer) the environment variables for the error handler will be set.

The location of the error log file is:

$ERRMON_LOGDIR/errorfile136574.log

The interpretation of the error is done by the error handler, it is centralized so that the operator has one place to look.

Alerts the user of serious error conditions. Currently an orange window appears when there is a fatal error condition. Text message instructs the operator what to do.

In addition to the visual alert there is a voice alert which states the problem.

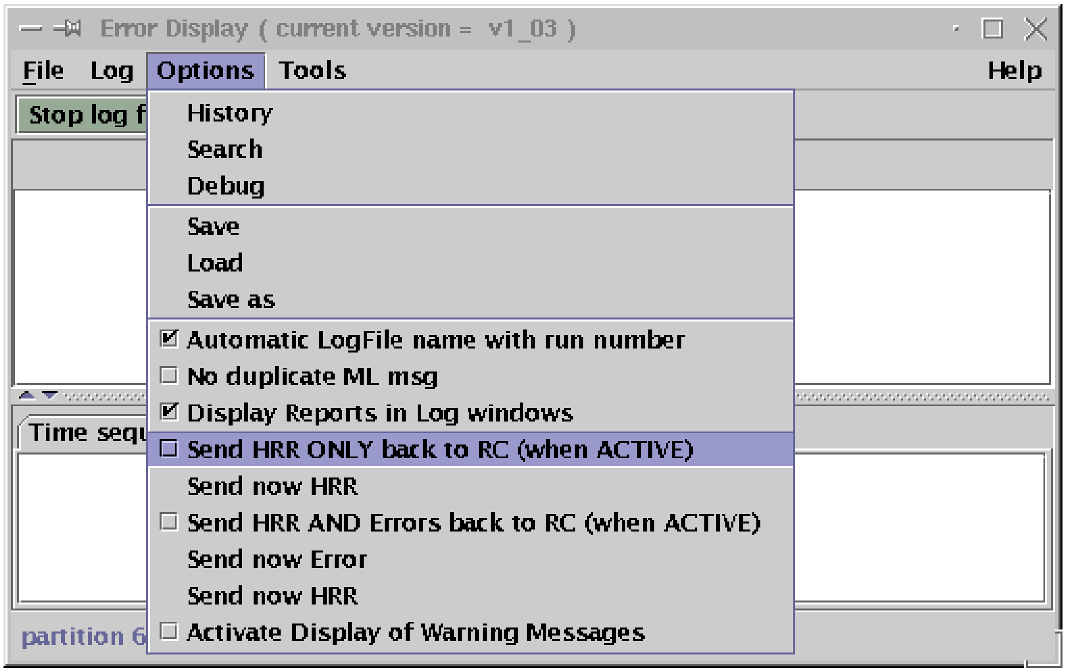
Can also be used to issue an automated Halt-Recover-Run sequence in the case of a Done or Busy timeout.

You should always be running with the automatic HRR enabled.

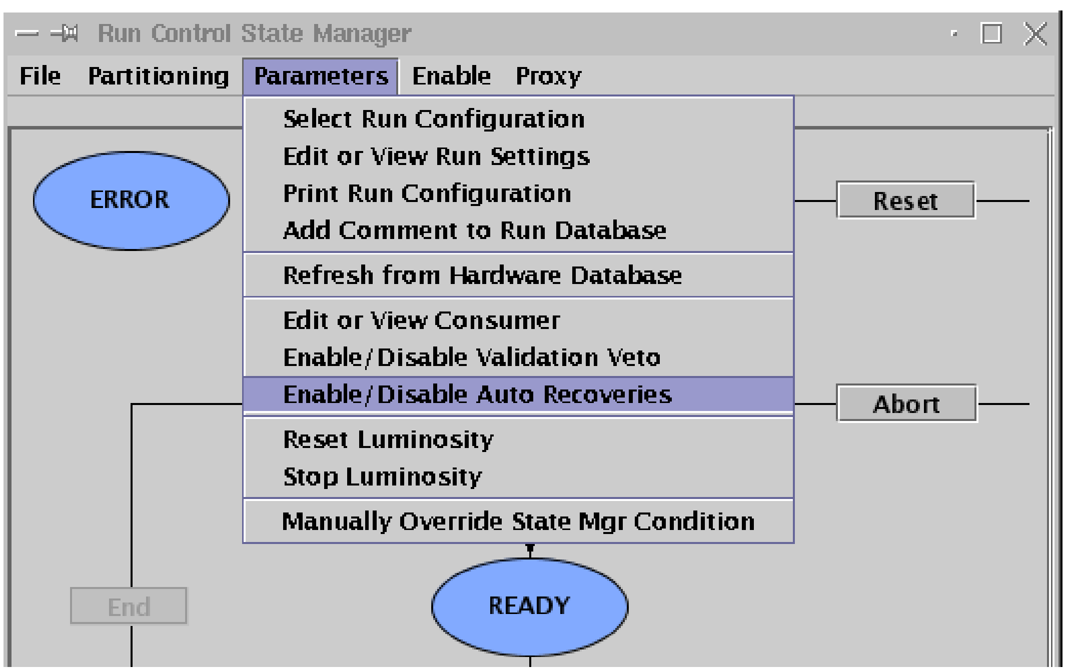
Report any problems with this feature in the shift log and send email to the RC email address:

`cdf-rc-support@fnal.gov`

You have to enable this option of sending commands from the error handler to RC.



and you have to enable RC to accept remote commands from the error handler.



Will build more intelligence into the error handler → expert system.

Level 3

Consists of a farm of dual processor PCs running Linux.

The raw data is complete and in the final data format when it gets to L3 and the offline reconstruction code is run to select events.

Reconstructed objects are added to the event record.

For example, using Edm_ObjectLister gives:

```
   129  LRIH_StorableBank       (   1: LRIH,    1, 0) RAW
   161  TFRD_StorableBank       (  10: TFRD,    1, 0) RAW
...
  7957  CalData                 (  47:    0,       0) L3
```

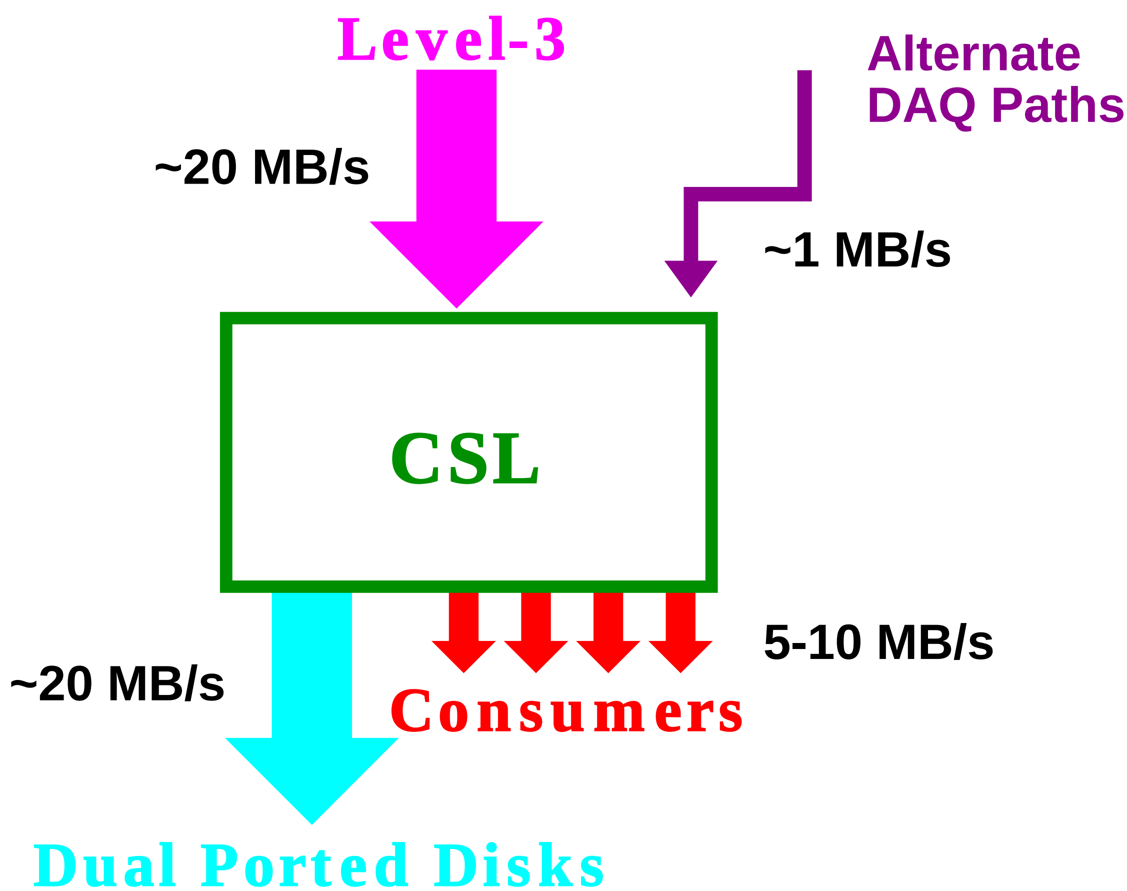
More details in a separate talk...

CSL

If an event passes a L3 trigger it is first sent to an output node then to the Consumer Server Logger (CSL).

The CSL distributes events to the various consumers which are used to check the quality of the data and the proper functioning of the trigger system.

The CSL writes data to disk in B0 separating it into different data streams based on the L3 trigger decision.



Data is copied from the disk buffers at B0 to the FCC disks then written to tape.

Important CDF DAQ Processes

Run control communicates with several key processes via so called proxy processes.

These processes are normally running but on rare occasions you may need to restart them.

In addition to these there are a number of other essential processes, for a summary see the "Important CDF DAQ Processes" link from the ace help page where you can find instructions on starting the processes.

SmartSockets b0dau30
Calibration Consumer Proxy b0dap62
Software EVB Proxy b0dap62
Resource Manager b0dap63
Consumer Monitor Proxy b0dap63
DBbroker Proxy b0dap63
L3Manager b0dap31
ACNET Monitor b0dap68
SVTSPYMON b0dap68
Consumer Server Logger b0dau32
Calibration Consumer b0dap60
Partition 14 Sender b0dap60

In addition there are

Consumer Disk Server b0dap65
Silicon Disk Server b0dap41

Consumers

Various consumers are used to check the quality of the data.

These are essentially AC++ modules compiled within the consumer framework used to monitor the quality of the data and the performance of the trigger.

Event Display
YMon
TrigMON
XMon
LumMon
Stage0
SiliMon
ObjectMon
BeamMon
L3RegionalMon
SVXMon
SVTMon
DAQMon

Used to identify hot channels (channels that are always on or are noisy), and dead regions (broken cables, high voltage problems...).

During shift operations there is a dedicated person (CO - Consumer Operator) assigned to look at the data quality.

Details in separate talk and at:

`http://www-b0.fnal.gov:8000/consumer/howto.html`

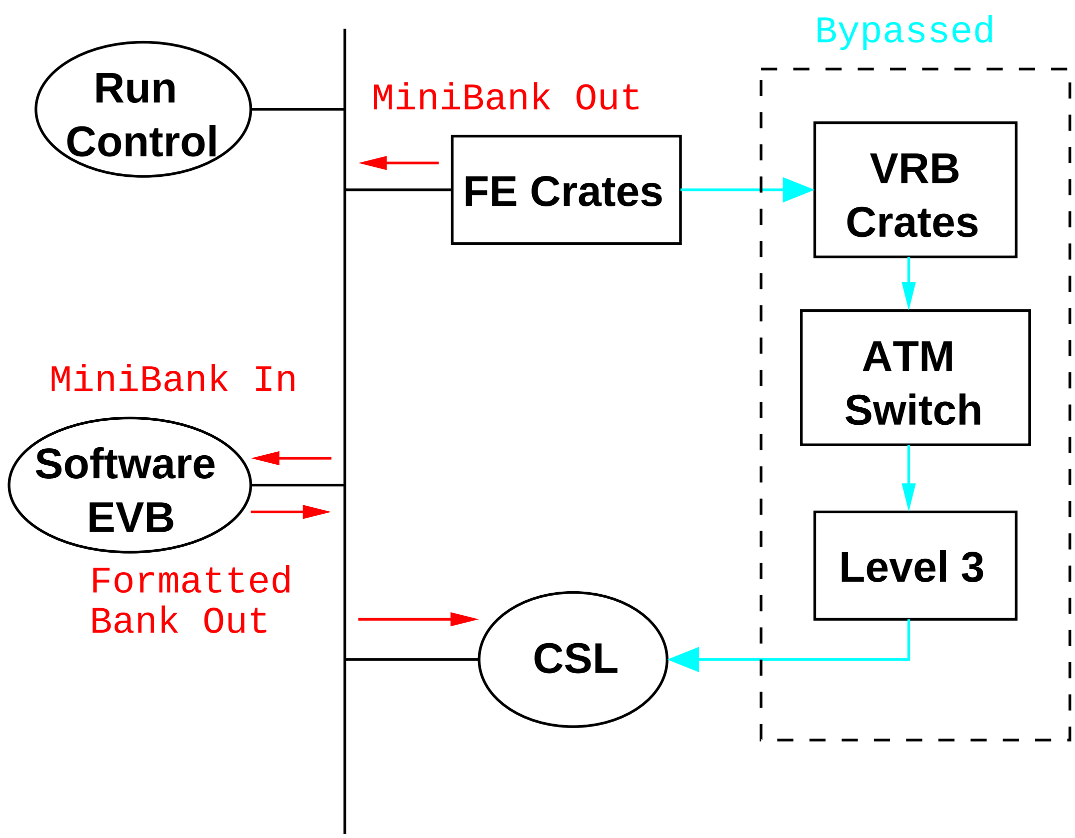
DAQMon Consumer

The DAQMon consumer plots the average readout time for each front end crate and the event size.

Noisy channels can sometimes lead to long readout times and large event fragments.

Software Event Builder

FE crates can send the mini banks over ethernet to a software client that collects the event fragments and reformats then into the final data format.



Events are sent to the CSL and can be distributed to the consumers or written to disk.

Used for debugging parts of the system and for calibrations.

Depending on how much data is being read out the rates can range from a few Hz to a few tens of Hz.

More details at:

`http://www-b0.fnal.gov:8000/ace2help/sevb/`

Calibrations

Calibrations for the different subsystems are also performed using Run Control.

The set of calibrations is part of the ace's duties.

Typically the software event builder is used for calibrations. This can accommodate larger event sizes.

Calibration data is sent to a *calibration consumer* which writes the results to a database.

Can view the results of the after being written into the database using DBANA.

Calibrations include:

Calorimeter - QIE, ShowerMax, LED and Xenon, Laser
CLC, COT, Silicon, Muon, TOF, BSC...

Details can be found at:

`http://www-b0.fnal.gov:8000/ace2help/ace_calibrations.html`

Also in a separate talk...

Dead Time

More details about the sources of deadtime and how to identify the source can be found at:

`http://www-b0.fnal.gov:8000/ace2help/deadtime.html`

The rate limit into the EVB is about 375 Hz. The logging rate limit of the CSL is 20 MB/s which corresponds to about 75 Hz.

Actual rates depend on the data volume, number of clients in the run and detector occupancy.

Normally the trigger tables are defined such that the dead time is less that 5%.

If the dead time is higher than $\sim 5\%$ then the source needs to be identified.

From the *DaqMon Rates and Deadtime* display you can see the total dead time of the system and the fraction from various sources.

The most common type of deadtime you will encounter are from "Busy" and "Readout".

Busy

This indicates that the VRB buffers are full and cannot accept more data.

Either the L3 accept rate is too high (faulty trigger) and we are limited by the 20 MB/s CSL rate, it is taking too long to process events at L3 or it is taking too long to read in the events or L3.

If the L3 display (one of the DAQ monitors) is "mostly green" this indicates that the processors are mostly occupied by trying to *output* the events to the CSL.

Check the CSL logging rate, if it is around 20 MB/s it may indicate that a L1 or L2 trigger is firing at too high of a rate.

One of the disks that the CSL buffers data onto is a slow disk and when writing to it we see logging rates of 15 MB/s.

A L3 display that is mostly "dark blue" indicates that the processors are busy *processing* the event. So far we have not been limited by the processing capacity of L3.

A L3 display that is mostly "light blue" indicates that the processors are busy *inputting* events. A BUSY for this case may occur if the event size is very large, for example noisy channels can lead to large events...

Readout

Readout deadtime occurs when the FE processors are taking too long to readout the event.

Many systems have a fixed data size, however for some the data volume increases with increasing luminosity.

For the TDCs the DSP processing time also increases with the number of "hits" for a channel.

$\rightarrow$ A typical source of readout deadtime is high occupancies for the TDCs which occurs when some channels are oscillating resulting in many "hits". The TDC DSP cannot process events fast enough...

One can identify the "bad guy" by using the RXPT monitor to see which crate shows up as the last to return DONE.

L2 Deadtime

L2 or Readout Deadtime

At a L1 accept rate of about 3.5 KHz and a L2 accept rate of about 250 Hz we have seen a dead time of about 2% due to a combination of "L2" and "L2 or Readout".

This is believed to be the result of the way buffers are managed by L2 which effectively reduces the number of available buffers from four to three.

Typical Warnings/Errors

Warning: COT Truncated Data

For very high occupancy events or when there is noise on a channel the data coming from one of the COT crates can be larger that what can be held in the VRB buffer.

In this case we truncate the date and you will get an error message of the type:

```
(MLE) b0cot14:5:37:25 AM->Runtime Error 1, Event 4793: data truncated
(MLE) b0cot02:5:54:23 AM->Runtime Error 2, Event 53148: data truncated
```

Warning: Bunch counter mismatch

Each front end card is checked that the BC is consistent. If there is a mismatch this warning will be sent.

For some crates this is a known problem and we do not try to go through a HRR sequence since the problem is cleared on the next event.

In other crates this problem is serious and with result in a done timeout. In this case we do issue a HRR to recover.

Warning: Reformatter Errors

Events with corrupted data fragments cannot be assembled into an event and are rejected by the reformatter process.

If the instantaneous rate of reformatter errors (measured over 30 seconds) is greater than 1% the error handler will pop up a warning message.

In this case follow the instructions on the window.

Reformatter errors are usually the result of corrupted silicon data and if they persist the offending wedge may have to be removed by an expert.

Error: Done Timeout

If an error is detected in the front end crate the process may not set the “done” bit generating a done timeout.

The run can be recovered by going through the Halt Recover Run sequence.

→ This type of error is detected by the error handler and an automatic HRR is issued.

Error: Busy Timeout

Busy timeout occurs if the VRB data buffer is not emptied out fast enough and the front end process cannot send data to the VRB.

This type of error can be triggered by several causes.

→ This type of error is detected by the error handler and an automatic HRR is issued.

Error: On Transition

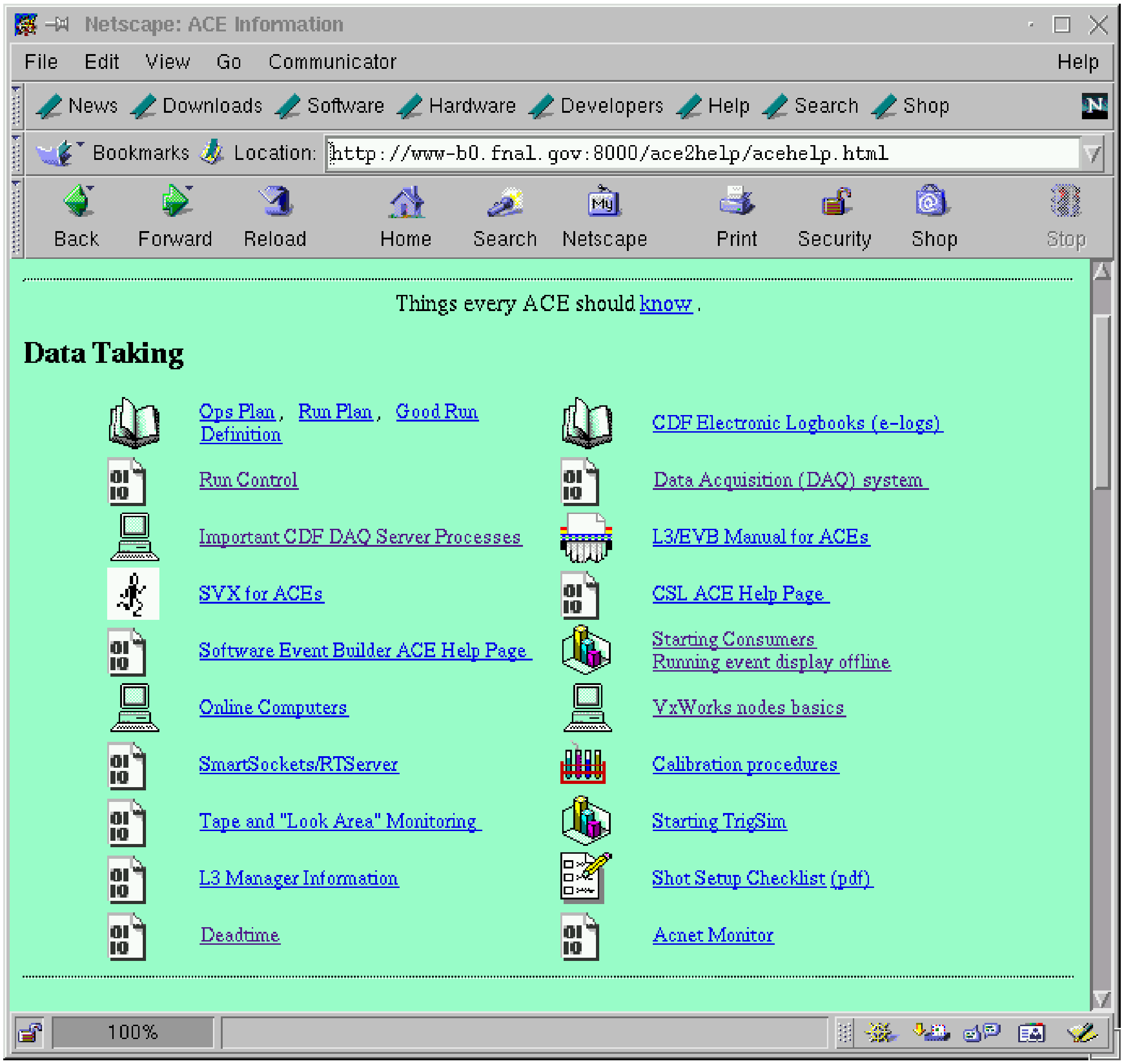
A transition can fail if there is a problem initializing the front end electronics.

One typical error during a transition is: “Error Initializing HDI”.

In this case one has to try the coldstart transition again.

Documentation

Many useful links to detailed documentation can be found on the Ace Help page: http://www-b0.fnal.gov:8000/ace2help/



Links to past Ace Training talks can be found there.

Many figures in this talk were lifted from previous talks...